

呼応する名詞の包含関係に着目した 助数詞オントロジーの自動構築と評価

白井 清昭[†]

徳永 健伸[‡]

[†]北陸先端科学技術大学院大学

[‡]東京工業大学

本論文では名詞と助数詞の呼応関係に基づく助数詞オントロジーの自動構築手法について述べる。まず、2つの助数詞に対し、それらと呼応関係にある2つの名詞集合に包含関係があれば、助数詞の間に上位-下位関係があると推測する。次に、獲得した上位-下位関係を基に、複数の木構造から構成される助数詞オントロジーを自動的に構築する。提案手法により、呼応関係にある名詞と助数詞の組、9,352組から助数詞オントロジーを自動構築する実験を行った。上位-下位関係がどれだけ信頼できるか、木構造が示唆する上位-下位関係の推移律がどの程度妥当であるか、などといった観点から助数詞オントロジーを評価し、妥当なオントロジーが得られたことを確認した。

Automatic Construction and Evaluation of Ontology for Japanese Classifiers based on Subsumption Relations of Agreeing Noun Sets

Kiyoaki SHIRAI[†]

Takenobu TOKUNAGA[‡]

[†]Japan Advanced Institute of Science and Technology

[‡]Tokyo Institute of Technology

This paper proposes a method for constructing an ontology of numerative classifiers based on noun-classifier agreement. Superordinate-subordinate relations are first extracted based on subsumption relations of noun sets corresponding to classifiers. An ontology is then automatically constructed using these extracted relations. We conducted an experiment to build an ontology from 9,352 pairs of noun-classifier pairs in Japanese. It was evaluated for the reliability of inferred superordinate-subordinate relations based on the generated ontology and the validity of transitivity of the relations in the ontology. We found the structure of the constructed ontology reasonable.

1 はじめに

日本語では名詞を数える際には一般に助数詞を必要とし、助数詞の種類も豊富である。さらに、例えば犬は「匹」では数えられるが「個」では数えられないように、ある名詞を数える際には特定の助数詞のみが使われるといった名詞と助数詞の呼応関係が存在する。計算機で助数詞を適切に取り扱うためには助数詞に関する知識の整備が必要不可欠である。

本研究は、助数詞に関する基礎的な知識として、日本語の助数詞オントロジーを自動的に構築することを目的とする。助数詞オントロジーとは、助数詞をその一般性に依じて体系的に整理した知識ベースで

ある。例えば、「頭」は比較的大型の動物を数えるときに使われるのに対し、「匹」は動物全般を数えるときに使われる。したがって、「匹」は「頭」よりも一般的な助数詞であるといえる。このような助数詞の一般性の違いを通常のオントロジーと同様に木構造で表現した知識体系を構築することが本研究の目標である。助数詞オントロジーは、自然言語解析や生成のための基礎的かつ有用な言語資源になりうる。また、日本語教育の面から、教育教材としての活用も期待できる。外国人、特に助数詞をあまり使わない欧州言語の使用者にとって、日本語の助数詞の使い方を覚えることは難しいとされている。助数詞のオント

